

Exploratory Data Analysis(EDA)



IPL

by

HARIKA KUDUMULA

Background

The Indian Premier League (IPL), is a professional Twenty20 cricket league in India, founded by the Board of Control for Cricket in India (BCCI) in 2008.

When?

April and May of every year

Who?

Teams representing 8 Indian cities and other states.

Seasons?

2008 - 2018 (eleven till date)

Data Profile

- Data Structure
 - Categorical and numerical data
 - Excel Sheet containing **696** records and **18** features
- Data Types
 - Integer
 - Object
- Missing values
 - Present

Features

```
id
season
city
date
team1
team2
toss_winner
toss_decision
result
dl_applied
winner
win_by_runs
win_by_wickets
player_of_match
venue
umpire1
umpire2
umpire3
```

Data Preprocessing

Data Normalization

- Date variable is considered as categorical variable and has different format for 2018 season data → **Convert to date format**
- ID column have discontinuous values (Season 2018) (in range of 7000 - 7953) → **Sorted by date & reassigned**
- Teams(14 distinct count), cities , venues had **alias names** → **Replaced aliases**

Missing Data and its imputation

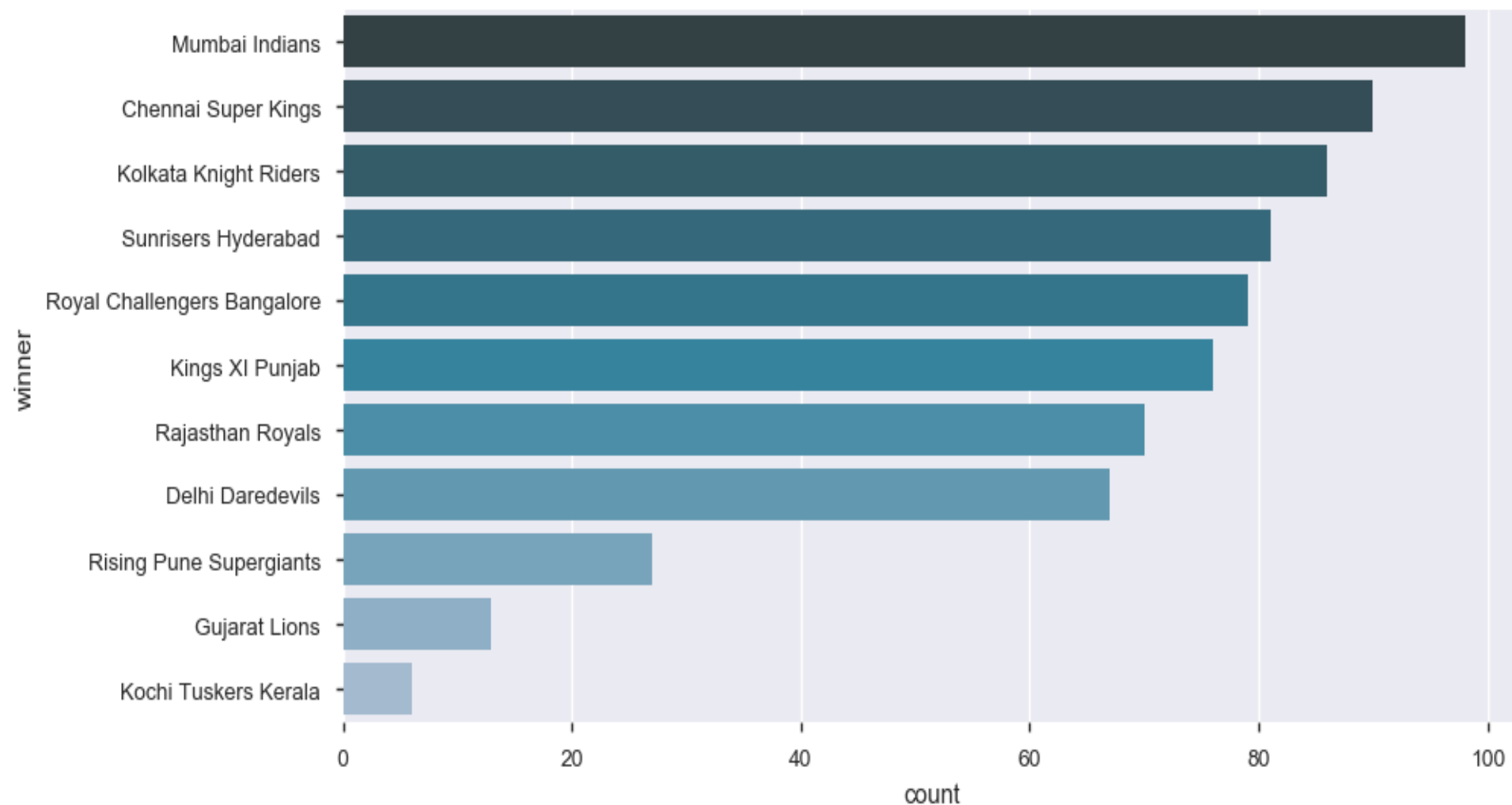
- Umpire 3 has 91.4% missing data → **Deleted umpire 3 data**
- City column has 7 missing values → **Find out by using venue details**
- Umpire 1,2 have 1 missing value → **Replaced with season MODE**

Problem statement

Understanding the IPL data by doing Exploratory Data Analysis and

- Finding patterns
- Determining relationships
- Preliminary understanding of important features in the data.

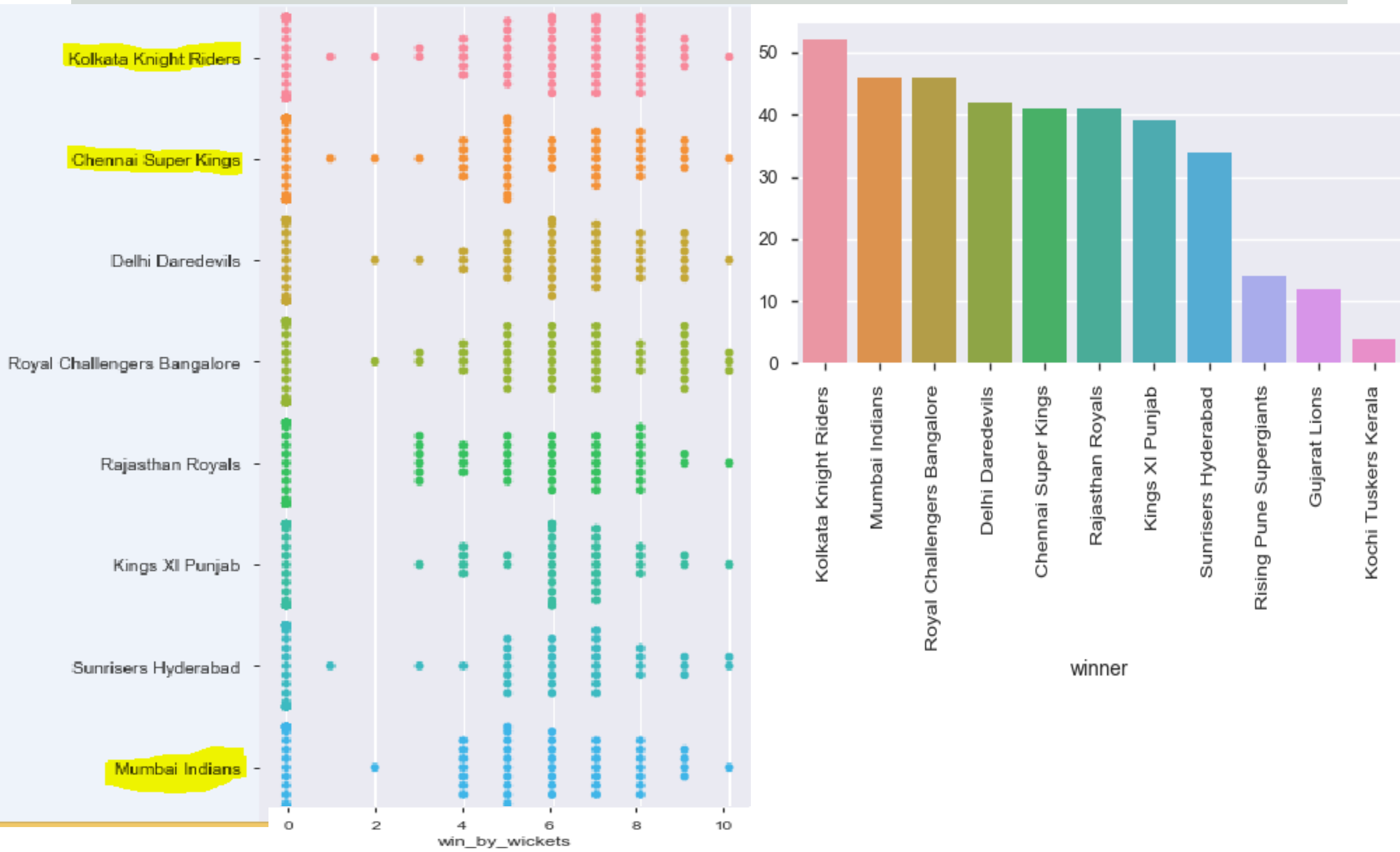
Best Teams



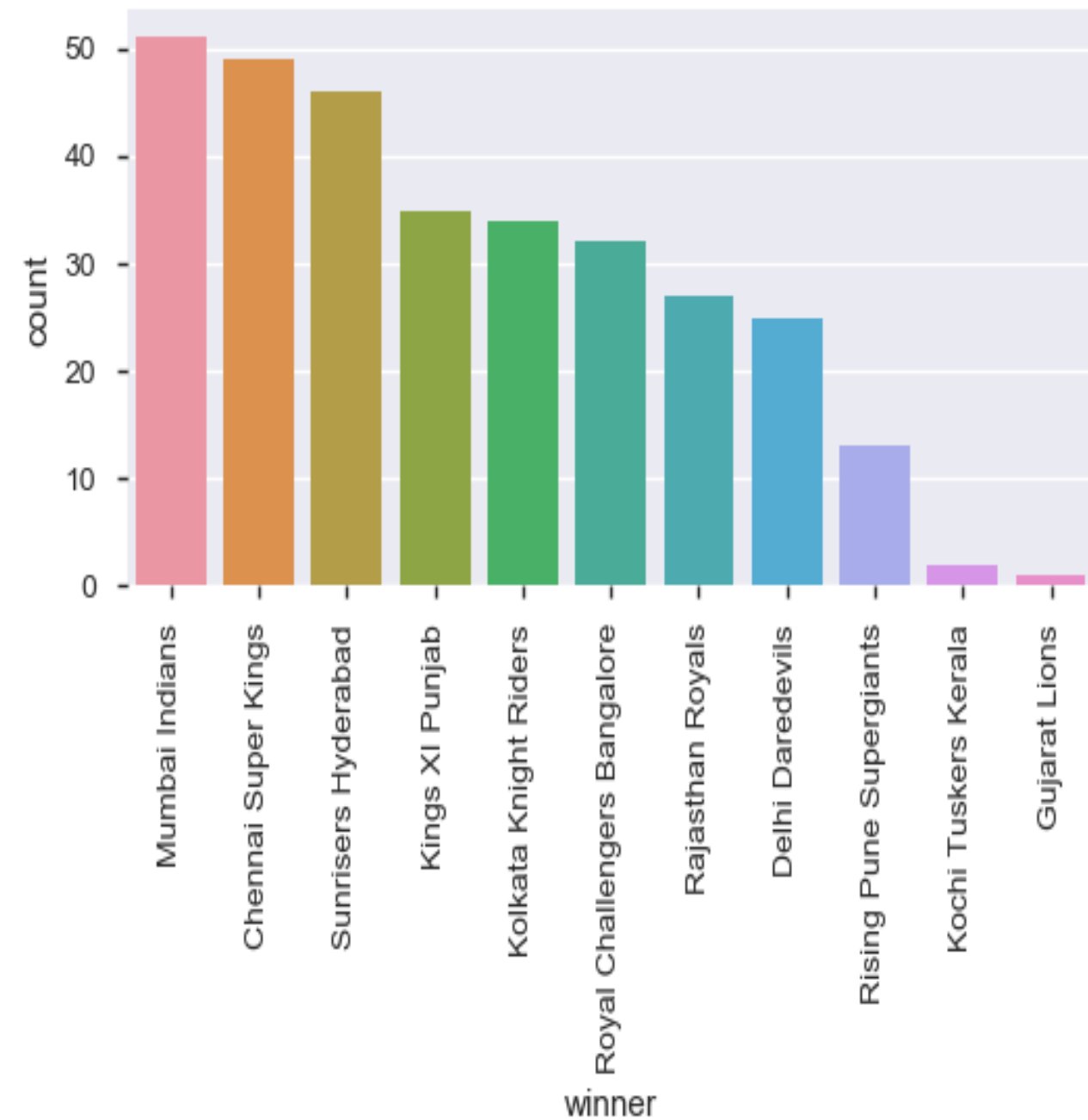
Most successful IPL Team –Mumbai Indians

season	winner
[2008]	[Rajasthan Royals]
[2009]	[Sunrisers Hyderabad]
[2010]	[Chennai Super Kings]
[2011]	[Chennai Super Kings]
[2012]	[Kolkata Knight Riders]
[2013]	[Mumbai Indians]
[2014]	[Kolkata Knight Riders]
[2015]	[Mumbai Indians]
[2016]	[Sunrisers Hyderabad]
[2017]	[Mumbai Indians]
[2018]	[Chennai Super Kings]

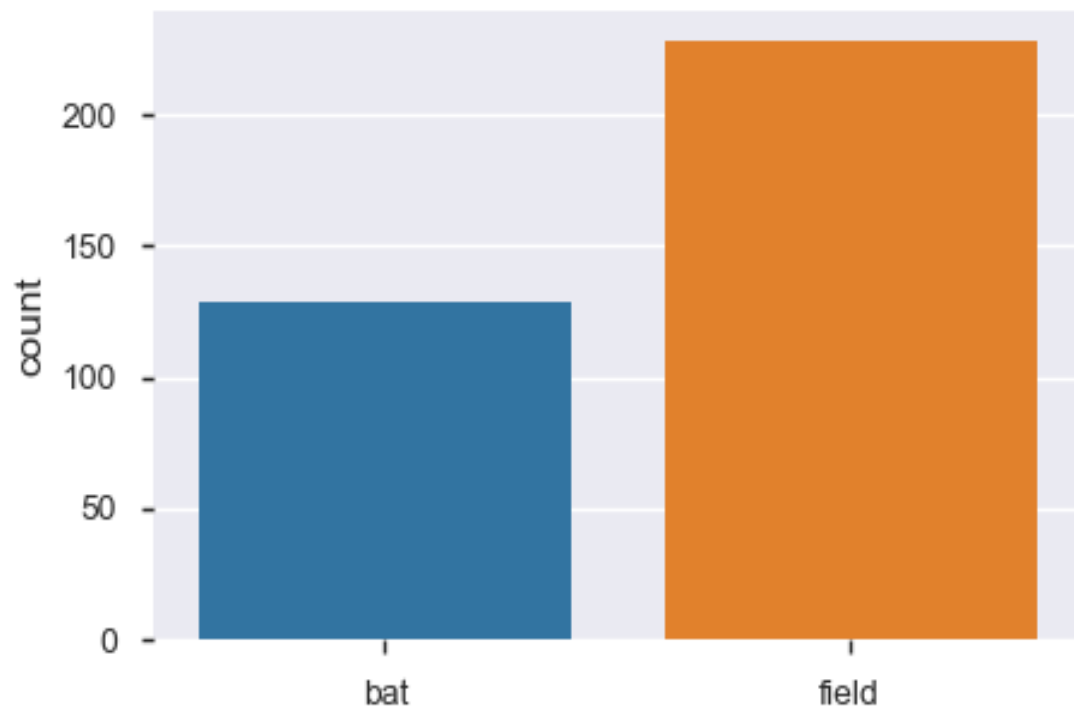
Best Chasing Teams



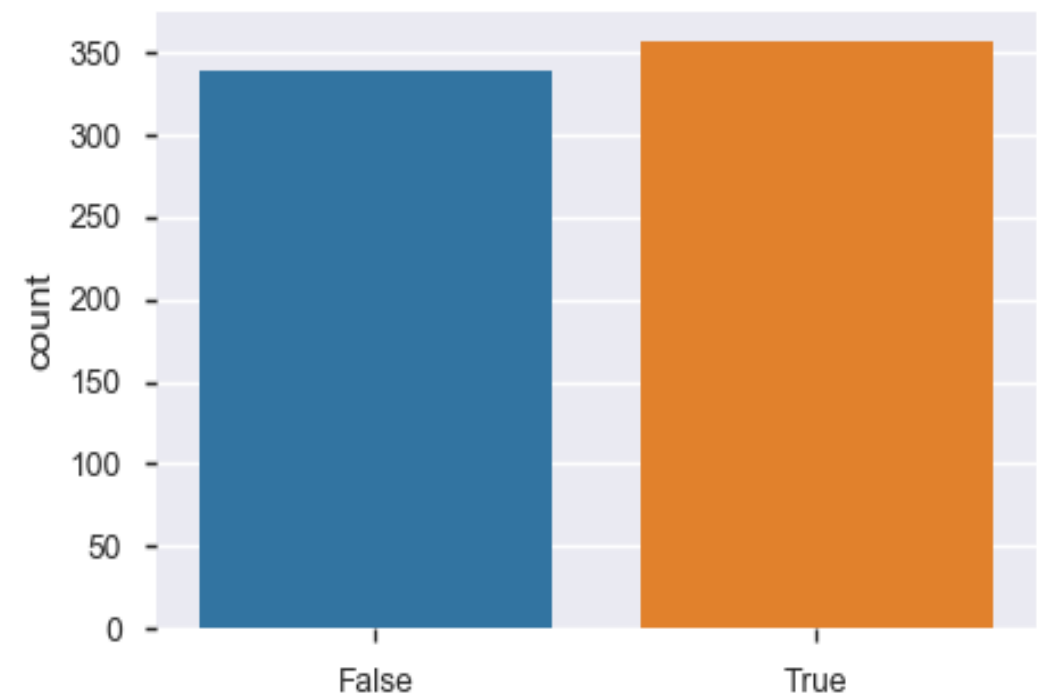
Best Defending Teams



Toss Decision



What toss decision helped toss winner to win the match ?

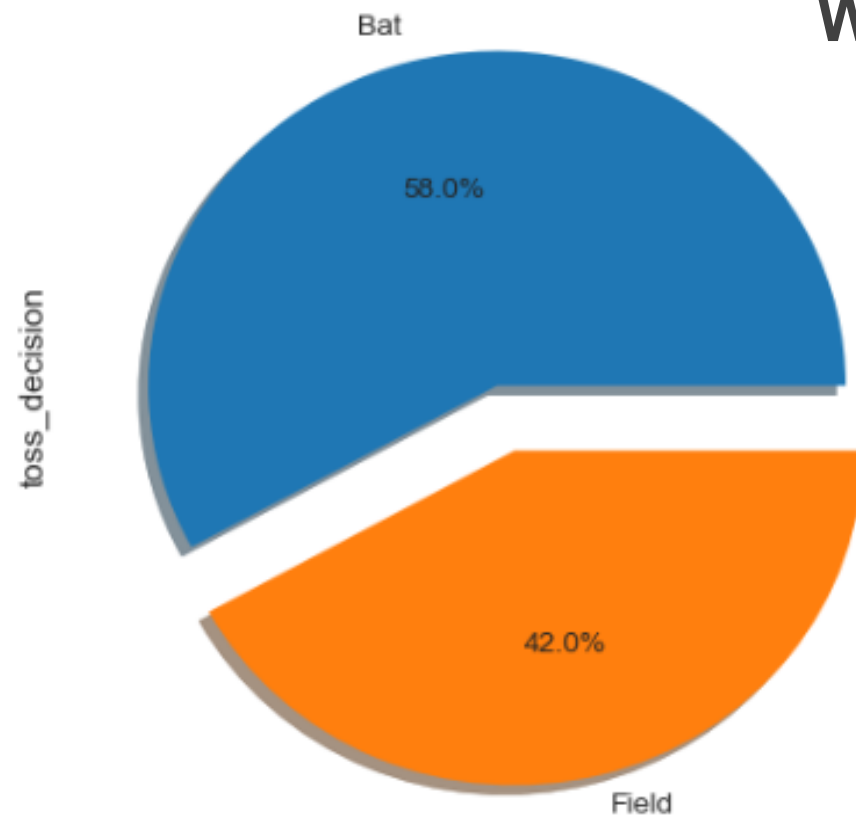


Has toss winning helped in match winning?

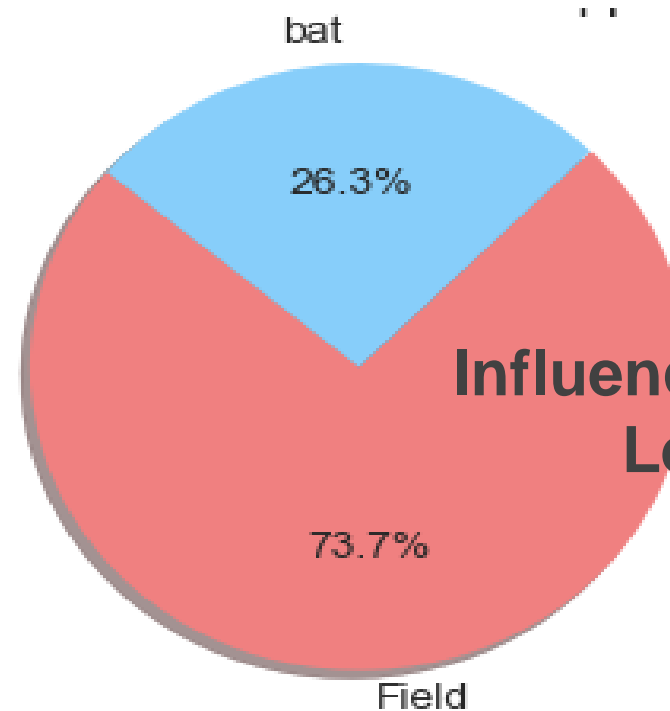
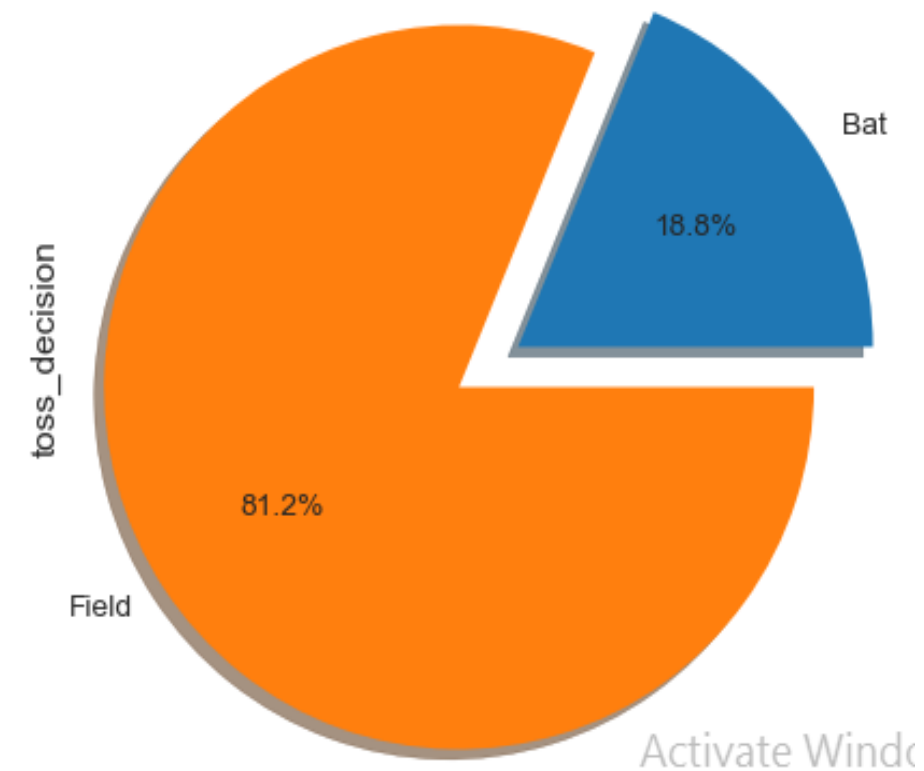
Toss Decision

What toss decision helped toss winner to win the match ?

Chennai Super Kings

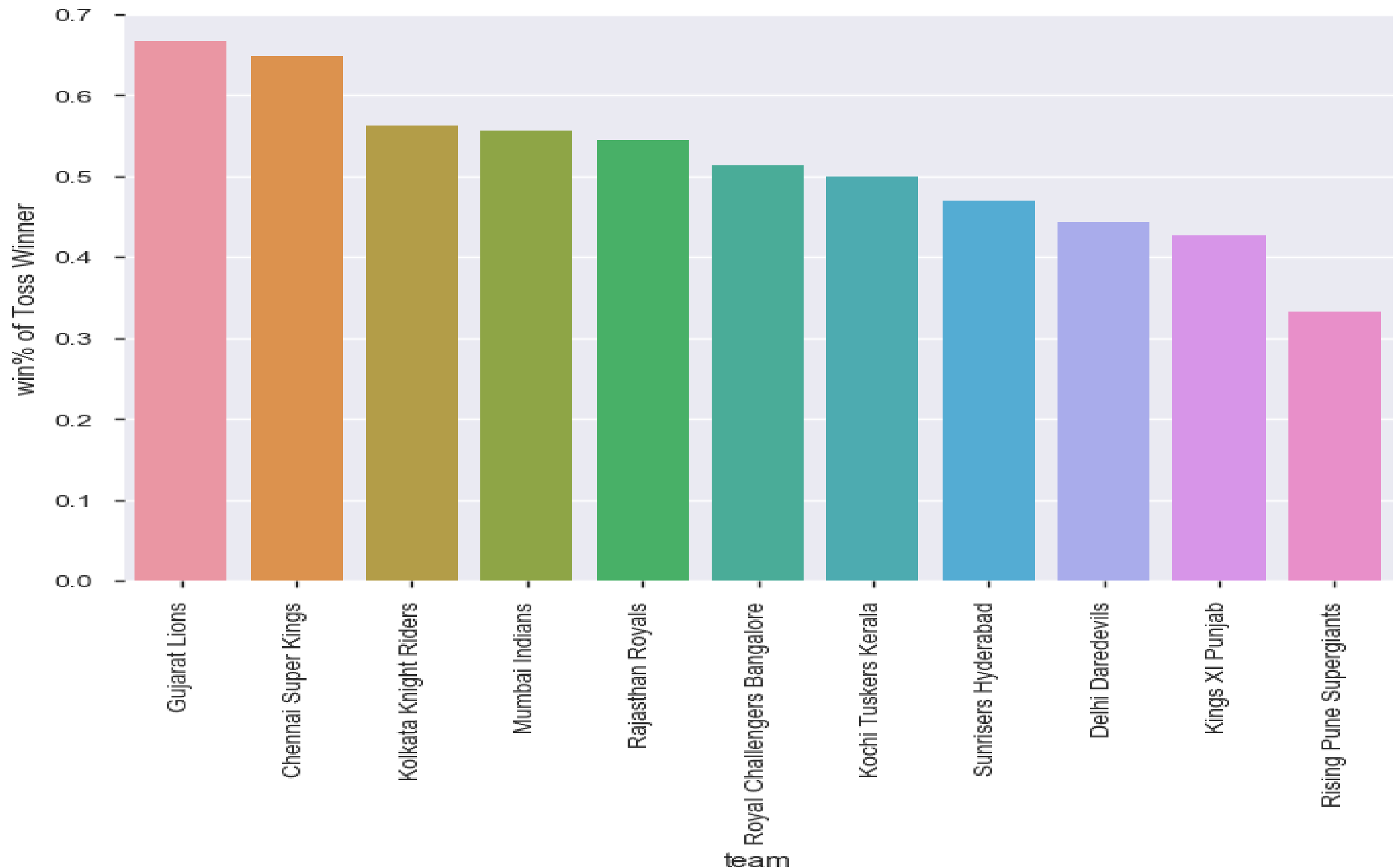


Kings XI Punjab

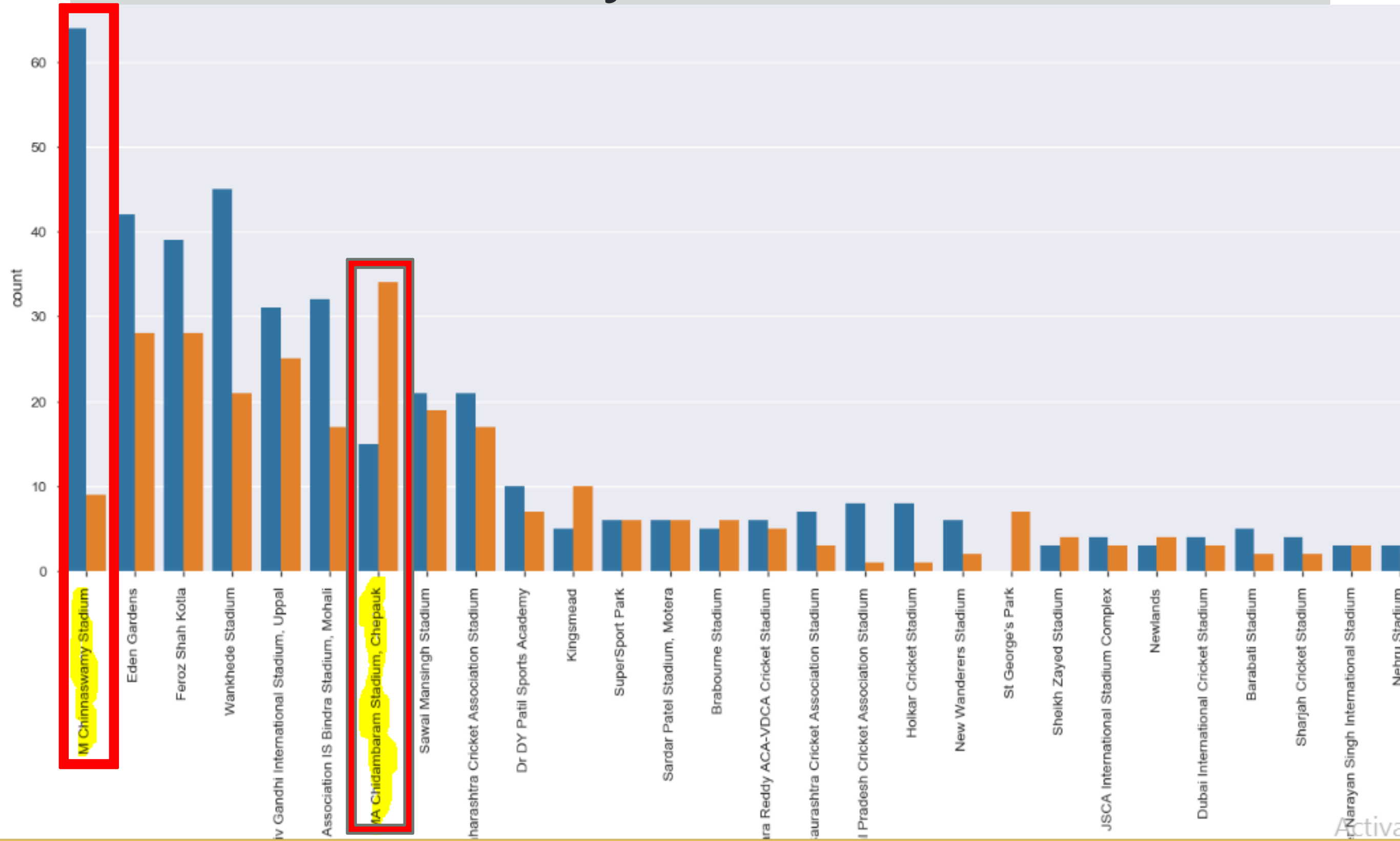


Influence of toss decision on Duckworth-Lewis method applied matches

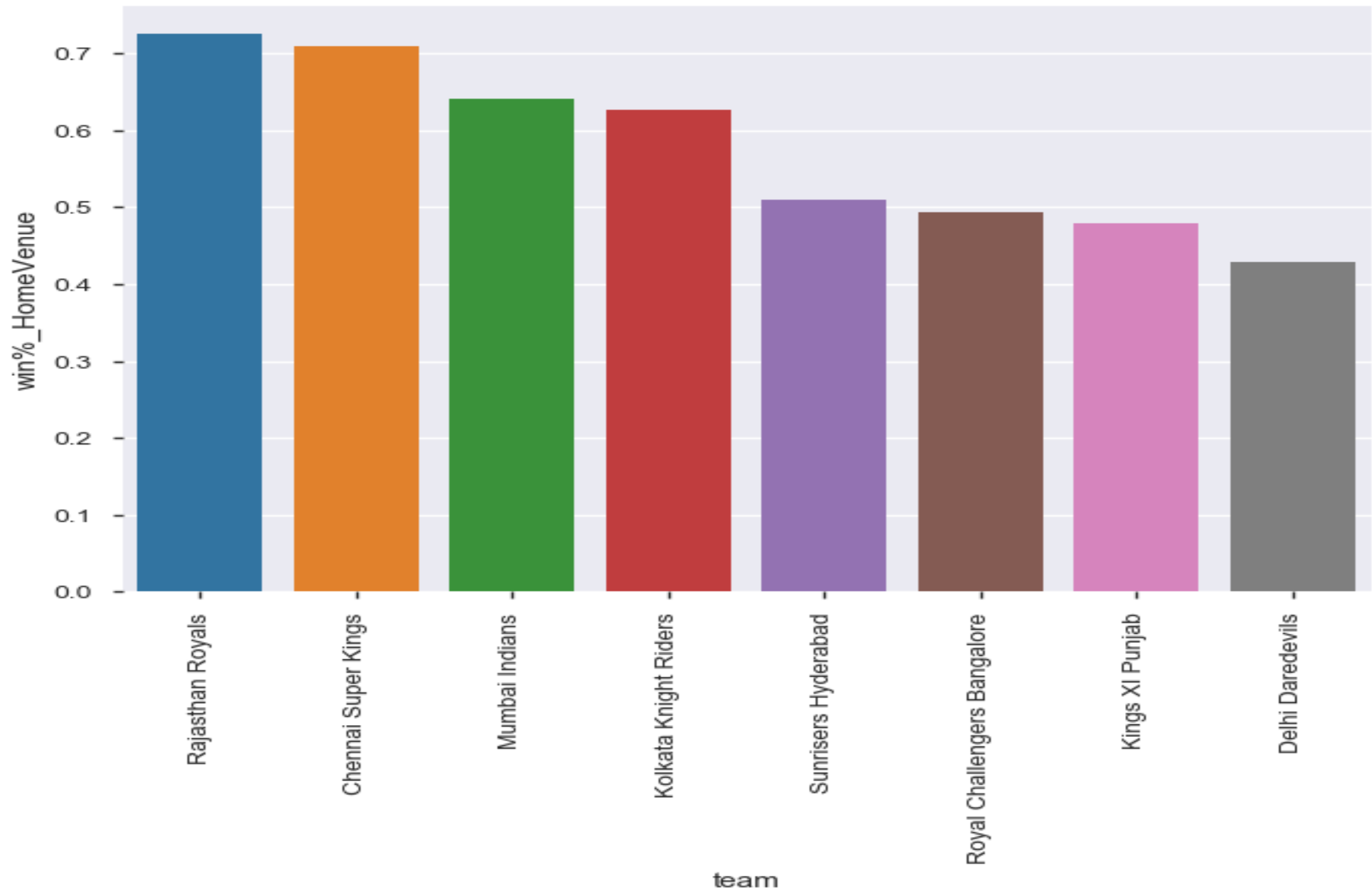
Win percentage of teams who won the toss



Toss Decision by venues



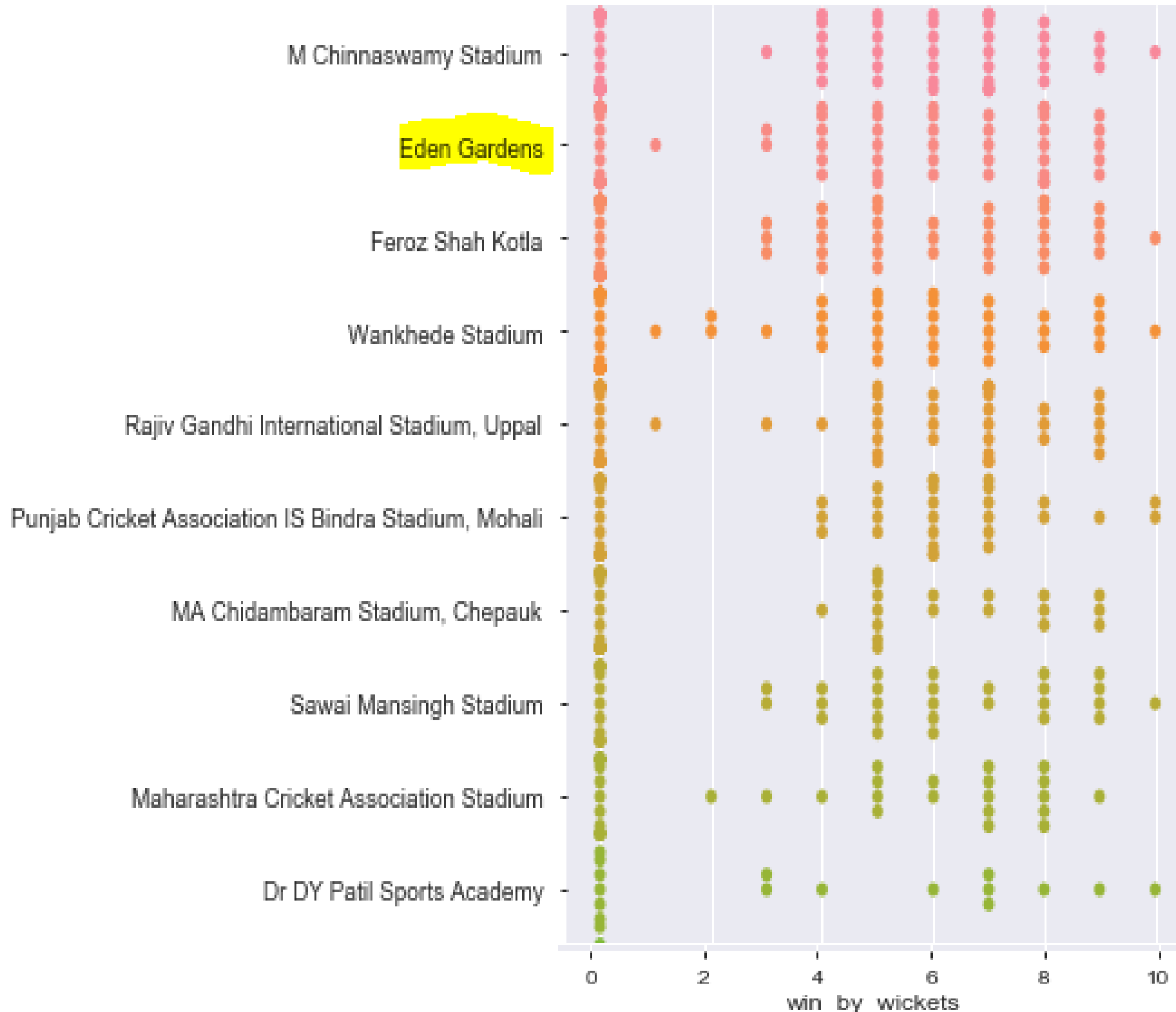
Influence of Home Venue on winning



Best defending venues



Best Chasing venues



Conclusion

- Mumbai Indians and Chennai Super Kings are both best chasing teams as well as defending teams, which justifies they being most title winners.
- Toss Decision of fielding first positively influences almost all IPL teams in all seasons.
- Home Venue has positive correlation with winning chances for major teams.
- Venue, toss decision has strong influence on the outcome of the match. Hence are important features for prediction models.

Actionable Items

- Venue, toss winner, toss decision have strong correlation with result of the match, hence these three variables should be included in any kind of prediction modelling on this data.
- As we have seen Mumbai Indians and Chennai Super Kings are the best teams, any marketing company can tap their reputation for promotion of their goods and services.